Computer Engineering and Applications Vol. 1, No. 2, December 2012

Combined Classifier for Face Recognition using Legendre Moments

D. Sridhar[1], Dr I.V. Murali Krishna[2]

[1]Balaji Institute of Technology and Science, Warangal (AP), India.

[2]Chief Consultant Offshore Programs in Asia, Jackson State University, USA

ABSTRAKSI

Dalam tulisan ini, metode baru gabungan Pengenalan Wajah berdasarkan pada momen Legendre dengan Analisis Diskriminan Linear dan Jaringan Syaraf Probabilistik diusulkan. Momen Legendre adalah invariants ortogonal dan skala sehingga mereka cocok untuk mewakili fitur dari gambar wajah. Metode yang diusulkan pengenalan Wajah terdiri dari tiga langkah, i) Fitur ekstraksi menggunakan momen Legendre ii) pengurangan dimensi menggunakan Analisis Discrminant Linear (LDA) dan iii) klasifikasi menggunakan Jaringan Syaraf Probabilistik (JSP). Analisis Diskriminan linier mencari petunjuk untuk diskriminasi maksimum kelas di samping pengurangan dimensi. Kombinasi momen Legendre dan Analisis Diskriminan Linear digunakan untuk meningkatkan kemampuan Analisis Diskriminan Linear ketika beberapa contoh gambar yang tersedia. Jaringan Syaraf Probabilistik memberikan klasifikasi cepat dan akurat gambar wajah. Evaluasi dilakukan pada dua basis wajah data. Basis data pertama dari 400 gambar wajah dari Laboratorium Penelitian Olivetty (ORLbasis data wajah, dan basisdata kedua siswa tiga belas diambil. Metode yang diusulkan memberikan tingkat pengenalan yang cepat dan lebih baik bila dibandingkan dengan pengklasifikasi lainnya.

Kata Kunci: Momen Legendre Ortogonal, Analisis Diskriminan Linear (LDA), Jaringan Syaraf Probabilistik, Pengurangan Dimensi, Ekstraksi Fitur, Tingkat Pengenalan.

ABSTRACT

In this paper, a new combined Face Recognition method based on Legendre moments with Linear Discriminant Analysis and Probabilistic Neural Network is proposed. The Legendre moments are orthogonal and scale invariants hence they are suitable for representing the features of the face images. The proposed face recognition method consists of three steps, i) Feature extraction using Legendre moments ii) Dimensionality reduction using Linear Discrminant Analysis (LDA) and iii) classification using Probabilistic Neural Network (PNN). Linear Discriminant Analysis searches the directions for maximum discrimination of classes in addition to dimensionality reduction. Combination of Legendre moments and Linear Discriminant Analysis is used for improving the capability of Linear Discriminant Analysis when few samples of images are available. Probabilistic Neural network gives fast and accurate classification of face images. Evaluation was performed on two face data bases. First database of 400 face images from Olivetty Research Laboratories (ORL) face database, and the second database of thirteen students are taken. The proposed method gives fast and better recognition rate when compared to other classifiers.

Keywords: Orthogonal Legendre Moments, Linear Discriminant Analysis (LDA), Probabilistic Neural Network ,Dimensionality Reduction, Feature Extraction, Recognition Rate.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

1. INTRODUCTION

Face classification [1] has a large number of applications including security, person verification, Internet communication and computer entertainment. In the face classification, a given face is compared with the faces stored in a face database in order to identify the person. But the wide range of variations in human face due to view point, pose, illumination and expression deteriorate the recognition performance of the Face recognition systems. Face recognition is one of the challenging problems in research, till now there is no unique solution for all face recognition problems [2]. But everyone accept that the face recognition system is good, if it has less computational complexity, good recognition performance and occupies less memory.

In any Face Recognition system Dimensionality Reduction and Feature Extraction are very important aspects. Face images though small in size are having large dimensionality this leads to very large computational time, complexity and memory occupation. The performance of any classifier mainly depends on high discriminatory features of the face images [2] – [4]. In our method we used Orthogonal Legendre moments and Linear Discriminant Analysis for dimensionality reduction and feature extraction.

Legendre moments preserve almost all the information of the image in few coefficients. Keeping these coefficients and ignoring the rest, we can reduce the dimensionality of the face image features. When dimensionality of the face images is high, Linear Discriminant Analysis is not applicable. To resolve this problem we combine the Legendre moments and Linear Discriminanat Analysis methods. By applying Legendre moments we get discriminatory features of the images of dimensionality 81, which are applied to Linear Discriminant Analysis. Linear Discriminant Analysis searches the directions for maximum discrimination of classes in addition to dimensionality reduction and finally produces image features of dimensionality of 39. Probabilistic Neural Network classifies the images based on their LDA features. The flow chart of the proposed method is shown in figure 1.

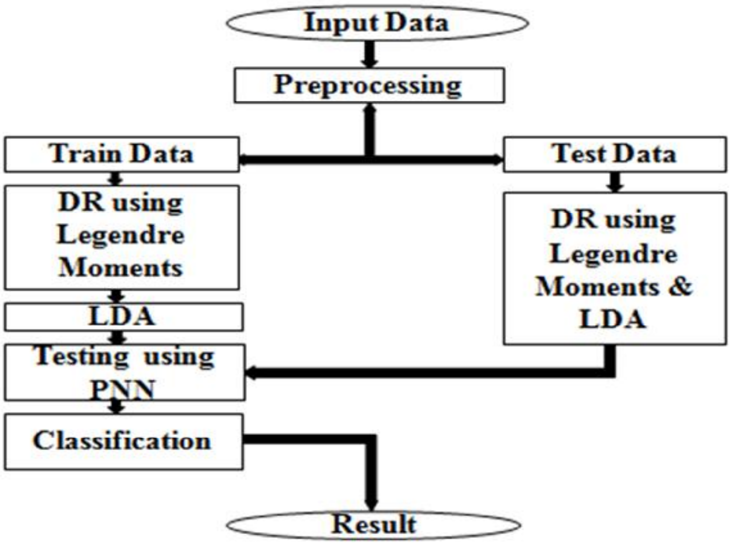


Figure 1. Flowchart of the proposed method

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

The rest of this paper is organized as follows: Section II discusses computation of Legendre moments for face images. Section III discusses the LDA. Section IV describes the PNN classifier. Section V shows experimental results, and discusses possible modifications and improvements to the system. Section VI presents concluding remarks.

2. LEGENDRE MOMENTS

Legendre moments were introduced by Teague [6],[7]. Legendre moments belong to the class of orthogonal moments and they were used in several pattern recognition applications. The definition of Legendre moments has a form of projection of the image intensity function into Legendre polynomials. The two-dimensional Legendre moments of order (p + q), with image intensity function f(x, y), are defined as

$$L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^{1}\int_{-1}^{1} P_p(x) P_q(y) f(x,y)\,dx\,dy \tag{1}$$

Where $x, y \in \{-1,1\}$

$P_p(x)$ is the p^{th} degree Legendre polynomial it is given by

$$P_p(x) = \sum_{k=0}^{p} \left\{ (-1)^{\frac{p-k}{2}} \frac{1}{2^p} \frac{(p+k)!\,x^k}{\left[\frac{p-k}{2}\right]!\left[\frac{p+k}{2}\right]!K!} \right\}_{p-k=even} \tag{2}$$

The recurrence relation of Legendre polynomials $P_p(x)$ is given as follows

$$P_p = \frac{(2p-1)P_{p-1}(x) - (p-1)P_{p-2}(x)}{P} \tag{3}$$

Where $P_0(x) = 1$, $P_1(x) = x$ and $p > 1$. Since the region of definition of Legendre polynomials is the interior of $\{-1,1\}$, a square image of N×N pixels with intensity function f(i, j), $0 \le i,\ j \le (N-1)$ is scaled in the region of $-1 < x, y < 1$. As a result of this equation (1) can now be expressed in discrete form as

$$L_{pq} = \lambda_{pq} \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_p(x_i) P_q(y_i) f(i,j) \tag{4}$$

Where the normalizing constant $\lambda_{pq} = \frac{(2p+1)(2q+1)}{N^2}$ (5)

x_i and y_i denote the normalized pixel coordinates in the range of $\{-1,1\}$, which are given by

$$x_i = \frac{2i}{N-1} - 1 \quad \text{and} \quad y_i = \frac{2i}{N-1} - 1 \tag{6}$$

3. LINEAR DISCRIMINANT ANALYSIS (LDA)

Linear discriminant analysis (LDA) tries to find the subspace that best discriminate different face classes in addition to dimensionality reduction [5], [8] – [14]. This is achieved by maximizing the ratio of the determinant of the between – class scattering matrix of the projected

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

samples to the determinant of the within – class scattering matrix. Within – class scattering matrix is defined as

$$S_W = \sum_{i=1}^{C} \sum_{x \in \{c_i\}} (x - m_i)(x - m_i)^T \tag{7}$$

Where **C** is the number of classes, $\mathbf{C_i}$ is a set of data within the i^{th} class, and m_i is the mean of the i^{th} class. The within class scatter matrix represents the degree of scatter within classes as a summation of covariance matrices of each class. A total scatter matrix $\mathbf{S_T}$ and a total mean $\mathbf{m}$ are defined as

$$S_T = \sum_x (x - m)(x - m)^T \tag{8}$$

and $$m = \frac{1}{n}\sum_x x = \frac{1}{n}\sum_{i=1}^{C} n_i m_i \tag{9}$$

Where n is the number of total samples and n_i is the number of samples within the i^{th} class. Then we get $S_T = S_W + \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T$ (10)

The second term in the above equation is defined as a between – class scatter matrix S_B, so that the total scattering matrix is the sum of the within – class scatter matrix and the between – class scatter matrix.

$$S_B = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T \tag{11}$$

and $$S_T = S_W + S_B$$

The between – class scatter matrix represents the degree of scatter between classes as a covariance matrix of means of each class. The projection of d – dimensional input samples onto r – dimensional space (r << d) is done by $y = W^T x$. We can obtain the transformation matrix W as one that maximizes the criterion function J(W) given as

$$J(W) = \arg\max_W \left| \frac{W^T S_B W}{W^T S_W W} \right| \tag{12}$$

The columns of optimal W are the generalized eigenvectors w_i that correspond to the largest Eigen values in $S_B W_i = \lambda_i S_W W_i$

4. PROBABILISTIC NEURAL NETWORK

Probabilistic Neural Network is a type of Radial Basis Function (RBF) network, which is suitable for pattern classification. The basic structure of a probabilistic neural network is shown

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

in figures 2. The fundamental architecture has three layers, an input layer, a pattern layer, and an output layer.

The pattern layer constitutes a neural implementation of a Bayes classifier, where the class dependent Probability Density Functions (PDF) are approximated using a Parzen estimator [15]-[18]. Parzen estimator determines the PDF by minimizing the expected risk in classifying the training set incorrectly. Using the Parzen estimator, the classification gets closer to the true underlying class density functions as the number of training samples increases.

The pattern layer consists of a processing element corresponding to each input vector in the training set. Each output class should consist of equal number of processing elements. Each processing element in the pattern layer is trained once. An element is trained to return a high output value when an input vector matches the training vector. In order to obtain more generalization a smoothing factor is included while training the network. The pattern layer classifies the input vectors based on competition, where only the highest match to an input vector wins and generates an output.

Compared to the feed forward back propagation network, training of the probabilistic neural network is much simpler. Since the probabilistic networks classify on the basis of Bayesian theory, it is essential to classify the input vectors into one of the two classes in a Bayesian optimal manner. The Bayes rule classifies an input vector belonging to class A as,

$$P_A C_A f_A(x) > P_B C_B f_B(x) \tag{13}$$

Where,
P_A - Priori probability of occurrence of patterns in class A
C_A - Cost associated with classifying vectors
$f_A(x)$ - Probability density function of class A

The PDF estimated using the Bayesian theory should be positive and integratable over all x and the result must be 1. The probabilistic neural net uses the following equation to estimate the probability density function given by,

$$f_A(x) = \frac{1}{(2\Pi)^{n/2}\sigma^n}\frac{1}{m_n}\sum_{i=1}^{m_A}\exp\left[-2\frac{(x-x_A)^r(x-x_{Ai})}{\sigma^2}\right] \tag{14}$$

Where
x_{Ai} - ith training pattern from class A
n - Dimension of the input vectors
σ - Smoothing parameter (corresponds to standard deviations of Guassian distribution)

The function $f_A(x)$ acts as an estimator as long as the parent density is smooth and continuous. $f_A(x)$ approaches the parent density function as the number of data points used for the estimation increases. The function $f_A(x)$ is a sum of Guassian distributions.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

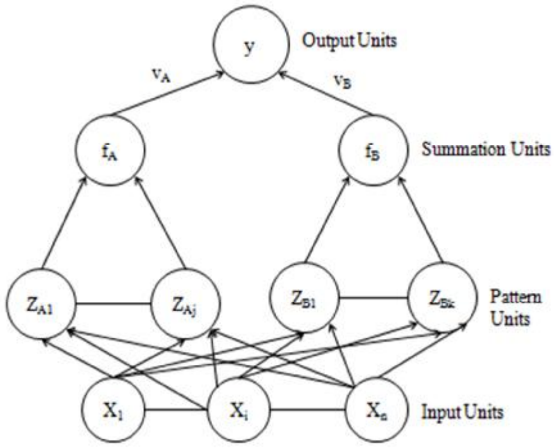


Figure 2. Architecture of Probabilistic Neural Network

5. NUMERICAL RESULTS

The proposed new face recognition method was applied on two different databases. The first database is AT&T (ORL) face database, it contains 40 subjects and each subject having 10 images of size 112 × 92 and the second is student database consists of thirteen students face images each with 5 images of different facial expressions, poses and background conditions. For simulations and proposed method evaluation Matlab is used on a PC with Intel(R) core (TM) 2 Duo CPU and 2 GB RAM. The obtained results for tow databases are given separately as follows.

5.1 AT&T (ORL) FACEDATABASE

Before doing the experiment, size of the face images are reduced to 10 × 10 using matlab. These reduced size images are used as inputs to the Legendre polynimial and it gives 81 features per image as output. But experimental results shows that 64 features per face image gives maximum recognition hence only 64 features per image are used for further use. For this database the graph drawn between Image size vs recognition rate is shown in figure 3 and the graph drawn between numbers of features per sample (Sample dimension) vs recognition rate is shown in figure 4. From these figures it is clear that Image size of 10 × 10 with sample dimension 20 is giving maximum recognition rate of 100%.

The feature vectors of the face images obtained at the output of the Legendre polynomials are given to the LDA as input. LDA produces 39 most discriminate features per image that leads better classification. The discriminant features of the face images produced by the LDA are given to the Probabilistic Neural Network for classification. Probabilistic Neural network (PNN) is a promising tool and gives fast and accurate classification of face images. The most important advantage of PNN is that training is easy and instantaneous [4]. Weights are not trained but assigned. Existing weights will never be alternated but only new vectors are inserted into weight

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

matrices when training. So it can be used in real – time. Since the training and running procedure can be implemented by matrix manipulation, the speed of PNN is very fast. In simulation three different combinations of training and test image samples are used as

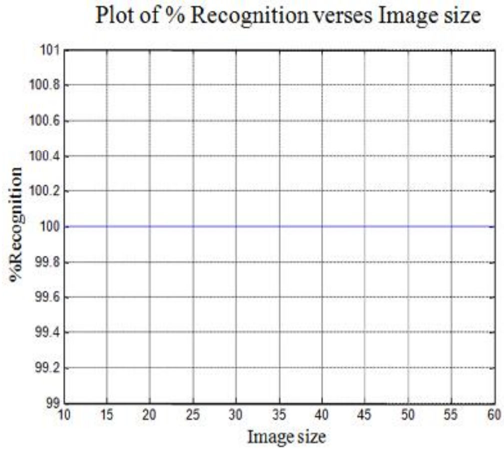


Figure 3. Plot of Recognition Rate verses image size

i) 280 training images and 120 test images
ii) 320 training images and 80 test images
iii) 360 training images and 40 test images

Simulation was done with the above three sets of data and the obtained results are summarized in the tabular form as shown in table 1.

No of features per image used in Legendre polynomials as input = 81
No of features per image used in LDA as input = 64
No of features per image used for classification = 20

For 320 training samples, the average training time taken is 4.6950 sec. For 80 test samples the average testing time taken is 11.466 sec that means 0.1433 sec time is taken for the testing of one sample. This is very small classification time when compare to any other Neural Network classification time. The test face image and recognized images are shown in figures 5(a) and 5(b).

Table 1. Recognition Rate for different Training and Test samples for AT&T (ORL) database

No of classes	No of Training Samples	Samples per class	No of Test Samples	Samples per class	Recognition Rate (Max)
40	280	7	120	3	99.1667
40	320	8	80	2	100
40	360	9	40	1	100

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

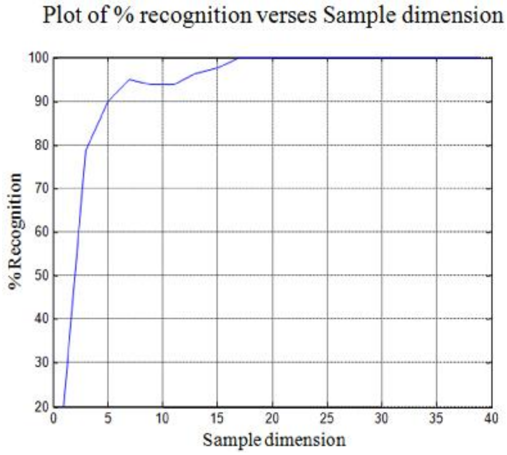
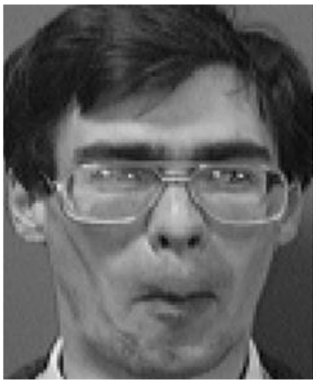


Figure 4. Plot of Recognition Rate versed Sample dimension



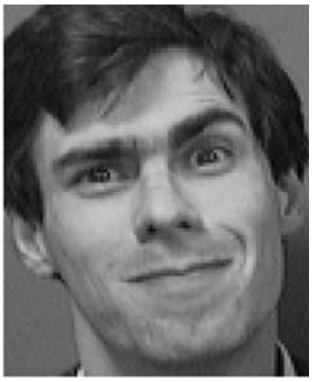


Figure 5(a). Test Face FIGURE 5(b). Recognized Face

5.2 STUDENTS DATABASE

This database consists of 13 classes and each class consists of 5 samples. One sample of each class is shown in figure 6. Before doing the experiment the size of the face images are reduced to 4×4 using matlab code. These reduced size images are used as inputs to the Legendre polynomials and it gives 16 features per image as output. But experimental results shows that 4 features per face image gives maximum recognition hence only 4 features per image are used for further use. The graph drawn between Image size vs recognition rate is shown in figure 7 and the graph drawn between number of features per sample (Sample dimension) vs recognition rate is shown in figure 8. From these figures it is clear that Image size of 4×4 with sample dimension 4 is giving maximum recognition rate of 100%.

The feature vectors of the face images obtained at the output of the Legendre polynomials are given to the LDA as input. LDA produces 12 most discriminate features per image that leads better classification. The discriminant features of the face images produced by the LDA are given to the Probabilistic Neural Network for classification. In simulations two different combinations of training and test image samples are used as

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

i) 52 training images and 13 test images
ii) 39 training images and 26test images

FIGURE 6. One sample from each Class of the Student database

Simulation was performed on the above two sets of data and the obtained results are summarized in the tabular form as shown in table 2.

No of features per image used in Legendre polynomials as input = 16

No of features per image used in LDA as input = 13

No of features per image used for classification = 4

For 52 training samples, the average training time taken is 1.1540 sec. For 13 test samples the average testing time taken is 2.4810 sec that means 0.1908 sec time is taken for the testing of one sample. This is very small classification time when compare to any other Neural Network classification time. The test face image and recognized images are shown in figures 9(a) and 9(b).

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

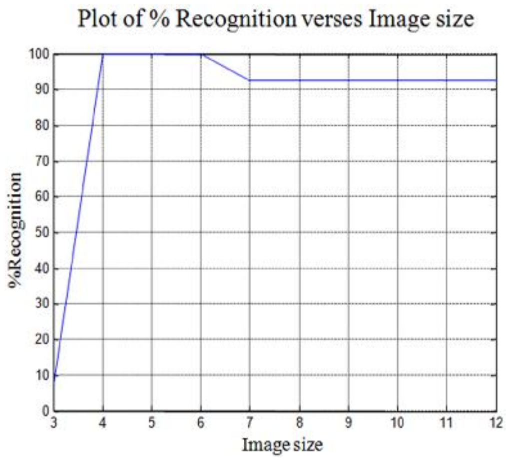


FIGURE 7. Plot of Recognition Rate verses image size

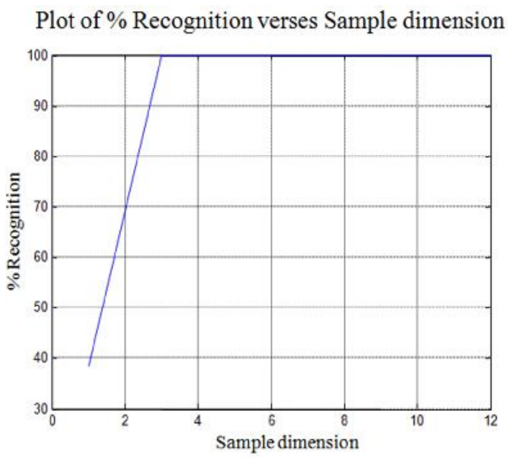


Figure 8. Plot of Recognition Rate verses Sample dimension

Figure 9(a). Test Face (b). Recognized Face

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Table 2. Recognition Rate for different Training and Test samples for Student database

NO of Classes	No of Training Samples	Samples per class	No of Test Samples	Samples per class	Recognition Rate (MAX)
13	39	3	26	2	92.3077 (24 out of 26)
13	52	4	13	1	100

6. CONCLUSION

In this paper, a new Face recognition method is presented. This new method is a combination of Legendre polynomials, Linear Discriminant Analysis and Probabilistic Neural Network. By using these algorithms an efficient face recognition method was constructed with maximum recognition rate of 100%. Simulation results using AT & T face database and student's database demonstrated the ability of the proposed method for optimal feature extraction and efficient face classification. The new face recognition algorithm can be used in many applications such as security systems.

The ability of our proposed face recognition method is demonstrated on the basis of obtained results on AT & T face database and student's database. For generalization, the proposed method should achieve 100% Recognition rate on other face databases and also on other combinations of training and test samples.

REFERENCES

1. A. Jain, R. Bolle, S. Pankanti Eds, *"BIOMETRIC – Personal Identification in Networked Society*". Kluwer Academic Publishers, Boston/Dordrecht/London,1999.
2. A.K. Jain, R.P.W. Duin, J. Mao, *Statistical pattern recognition: a review,* IEEE Trans. Pattern Anal. Mach.Iintell. 22 (2000) 4 – 37.
3. K. Fukunaga, *Introduction to Statistical pattern recognition*, II nd Eddition. Academic Press, New York, 1990.
4. N. Kwak, and C.H. Choi, *"Input Feature Selection for classification problem",* IEEE Transactions on Neural Networks, 13(1), 143-159,2002.
5. R. O. Duda, P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
6. Thawar Arif, Zyad Shaaban, Lala Krekor, Sami Baba, Object Classificatio Via Geometrical, Zernike and Legendre Moments, Journal of Theoretical and Applied Information Technology, 2005 – 2009, JATIT.
7. Khalid M. Hosny, Robust Templet Matching using Orthogonal Legendre Moment Invariants, Journal of computer science 6(10): 1083-1087,2010, ISSN 1549-3636.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

8. H. Jain, P.C. Yuen, C.Wen-Sheng, *A novel subspace LDA Algorithm for recognition of face images with illumination and pose variation,* Proceedings of the International Conference on Mission Learning and Cibernetics, Vol. 6, 2004,pp. 3589-3594.
9. W. zhao, R. Chellappa, N. Namdhakumar, *Emperical performance analysis of linear discriminant classifiers*, Proceeding of the IEEE Conference on Computer Vision and Pattern Recognitio, 1998,pp, 164-169.
10. W. zhao, R. Chellappa, P. J. Phillips, *Subspace linear discriminant Analysis for face recognition,* Technical Report CAR – TR -914, Center for Automation Research, University of Maryland, College Park,1999.
11. W. Xiaogang, T. Xiaoou, *A unified framework for subspace face recognition,* IEEE Trans. Pattern Anal. Math Intell. 26 (9) (2004) 1222 – 1228.
12. Alok Sharma, Kuldip K. Paliwal, Godfrey C. Onwubolu, *Ckass – dependent PCA, MDC AND LDA: A combined classifier fir pattern classification*, 2006 Pattern Recognition Society. Published by Elsevier Ltd. Doi:10,1016/j.patcog.2006.02.001.
13. Lu, K, Plataniotis, K.N. & Venetsanopoulos, A.N. (2003) *Face Recognition Using LDA – based algorithm,* IEEE trans. Neural Network, vol. 14, No 1, pp. 195- 199, January 2003.
14. D. Q. Dai and P. C. Yuen, *" Face recognition by regularized discriminant analysis"*, IEEE Trans. Vol. 37, no. 4, pp. 1080 – 1085, Aug,2007.
15. Dr.S.Annadurai, A.Saradha. " *Face recognition using Legendre Moments"*, Indian Conference on Computer Vision, Graphics and Image Processing 2004/proceedings-final.pdf.
16. D.F. Specht, *"Probabilistic Neural Networks for classification, mapping, or associative memory",* Proceedings of IEEE International Conference on Neural Networks, Vol 1, IEEE Press, New York, pp. 525-532, june 1988.
17. D.F. Specht, *"Probabilistic Neural Networks"* Neural Networks, Vol. 3, No 1,pp.109-118,1990.
18. Mohd Fauzi Othman and Mohd Ariffanan Mohd Basri, "*Probabilistic Neural Network For Brain Tumor Classification",* Second International Conference on Intelligent Systems, Modeling and Simulation, IEEE Computer Society, 2011, DOI 10.1109/ISMS. 2011.32.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)